(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 July 2003 (24.07.2003)

PCT

(10) International Publication Number
WO 03/061285 A2

(51) International Patent Classification[7]: H04N 7/16

(21) International Application Number: PCT/GB02/05908

(22) International Filing Date:
23 December 2002 (23.12.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0130936.8 24 December 2001 (24.12.2001) GB
0204323.0 25 February 2002 (25.02.2002) GB

(71) Applicant *(for all designated States except US)*: **SCIENTIFIC GENERICS LIMITED** [GB/GB]; Harston Mill, Harston, Cambridgeshire CB2 5GG (GB).

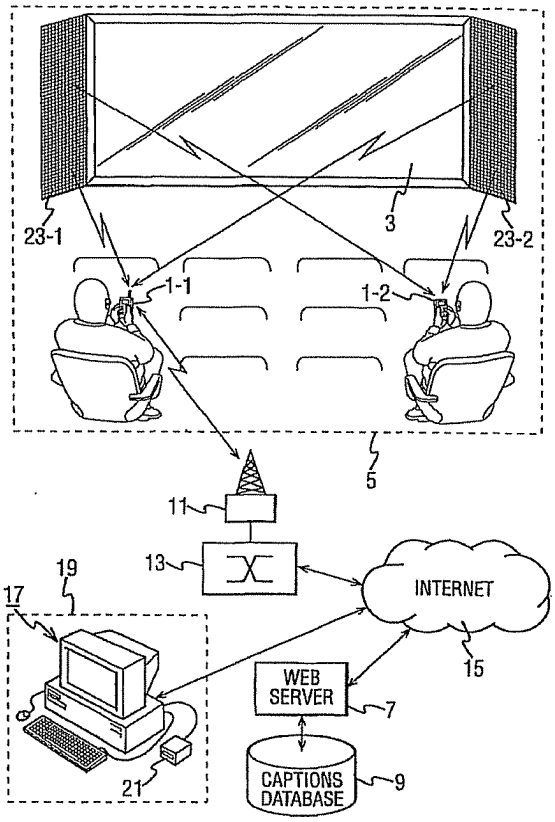
(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **HOSKING, Ian, Michael** [GB/GB]; Scientific Generics Limited, Harston Mill, Harston, Cambridgeshire CB2 5GG (GB). **JONES, Aled, Wynne** [GB/GB]; Scientific Generics Limited, Harston Mill, Harston, Cambridgeshire CB2 5GG (GB). **REYNOLDS, Michael, Raymond** [GB/GB]; Scientific Generics Limited, Harston Mill, Harston, Cambridgeshire CB2 5GG (GB). **KELLY, Peter, John** [GB/GB]; Scientific Generics Limited, Harston Mill, Harston, Cambridgeshire CB2 5GG (GB). **EMMENS, Michael, David** [GB/GB]; Scientific Generics Limited, Harston Mill, Harston, Cambridgeshire CB2 5GG (GB).

(74) Agents: **BERESFORD, Keith, Denis, Lewis** et al.; Beresford & Co., 2-5 Warwick Court, High Holborn, London WC1R 5DH (GB).

[Continued on next page]

(54) Title: CAPTIONING SYSTEM



(57) Abstract: A captioning system is provided for providing captions for audio and/or video presentations. The captioning system can be used to provide text captions or audio descriptions of a video presentation. A user device is provided for the captioning system having a receiver operable to receive the captions together with synchronisation information and a caption output circuit which is operable to output the captions at the appropriate timings defined by the synchronisation information. The user device is preferably a portable hand-held device such as a mobile telephone, PDA or the like.

WO 03/061285 A2

(81) **Designated States** *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

CAPTIONING SYSTEM

The present invention relates to a system and method and parts thereof for providing captions for audio or video or multi-media presentations. The invention has particular though not exclusive relevance to the provision of such a captioning system to facilitate the enjoyment of the audio, video or multimedia presentation by people with sensory disabilities.

A significant proportion of the population with hearing difficulties benefit from captions (in the form of text) on video images such as TV broadcasts, video tapes, DVD and films. There are currently two types of captioning systems available for video images - on-screen caption systems and off-screen caption systems. In on-screen caption systems, the caption text is displayed on-screen and it obscures part of the image. This presents a particular problem with cinema where there is a reluctance for this to happen with general audiences. In the off-screen caption system, the text is displayed on a separate screen. Whilst this overcomes some of the problems associated with the on-screen caption system, this solution adds additional cost and complexity and currently has had poor takeup in cinemas for this reason.

In addition to text captioning systems for people with hearing difficulties, there are also captioning systems which provide audio captions for people with impaired eyesight. In this type of audio captioning system, an audio description of what is being displayed is provided to the user in a similar way to the way in which sub-

titles are provided for the hard of hearing.

One aim of the present invention is to provide an alternative captioning system for the hard of hearing or an alternative captioning system for those with impaired eyesight. The captioning system can also be used by those without impaired hearing or eyesight, for example, to provide different language captions or the lyrics for songs.

According to one aspect, the present invention provides a captioning system comprising: a caption store for storing one or more sets of captions each being associated with one or more presentations and each set comprising at least one caption for playout at different timings during the associated presentation; and a user device having: (i) a memory for receiving and storing at least one set of captions for a presentation from the caption store; (ii) a receiver operable to receive synchronisation information defining the timing during the presentation at which each caption in the received set of captions is to be output to the user; and (iii) a caption output circuit operable to output to the associated user, the captions in the received set of captions at the timings defined by the synchronisation information.

In one embodiment, the captions are text captions which are output to the user on a display associated with the user device. In another embodiment, the captions are audio signals which are output to the user as acoustic

signals via a loudspeaker or earphone. The captioning system can be used, for example in cinemas, to provide captions to people with sensory disabilities to facilitate their understanding and enjoyment of, for example, films or other multimedia presentations.

The user device is preferably a portable hand-held device such as a mobile telephone or personal digital assistant, as there are small and lightweight and most users have access to them. The use of such a portable computing device is also preferred since it is easy to adapt the device to operate in the above manner by providing the device with appropriate software.

The caption store may be located in a remote server in which case the user device is preferably a mobile telephone (or a PDA having wireless connectivity) as this allows for the direct connection between the user device and the remote server. Alternatively, the caption store may be a kiosk at the venue at which the presentation is to be made, in which case the user can download the captions and synchronisation information when they arrive. Alternatively, the caption store may simply be a memory card or smart-card which the user can insert into their user device in order to obtain the set of captions for the presentation together with the synchronisation information.

According to another aspect, the present invention provides a method of manufacturing a computer readable medium storing caption data and synchronisation data for

use in a captioning system, the method comprising: providing a computer readable medium; providing a set of captions that is associated with a presentation which comprises a plurality of captions for playout at different timings during the associated presentation; providing synchronisation information defining the timing during the presentation at which each caption in the set of captions is to be output to a user; receiving a computer readable medium; recording computer readable data defining said set of captions and said synchronisation information on said computer readable medium; and outputting the computer readable medium having the recorded caption and synchronisation data thereon.

Exemplary embodiments of the present invention will now be described with reference to the accompanying drawings, in which:

Figure 1 is a schematic overview of a captioning system embodying the present invention;

Figure 2a is a schematic block diagram illustrating the main components of a user telephone that is used in the captioning system shown in Figure 1;

Figure 2b is a table representing the captions in a caption file downloaded to the telephone shown in Figure 2a from the remote web server shown in Figure 1;

Figure 2c is a representation of a synchronisation file

downloaded to the mobile telephone shown in Figure 2a from the remote web server shown in Figure 1;

Figure 2d is a timing diagram illustrating the timing of synchronisation signals and illustrating timing windows during which the mobile telephone processes an audio signal from a microphone thereof;

Figure 2e is a signal diagram illustrating an exemplary audio signal received by a microphone of the telephone shown in Figure 2a and the signature stream generated by a signature extractor forming part of the mobile telephone;

Figure 2f illustrates an output from a correlator forming part of the mobile telephone shown in Figure 2a, which is used to synchronise the display of captions to the user with the film being watched;

Figure 2g schematically illustrates a screen shot from the telephone illustrated in Figure 2a showing an example caption that is displayed to the user;

Figure 3 is a schematic block diagram illustrating the main components of the remote web server forming part of the captioning system shown in Figure 1;

Figure 4 is a schematic block diagram illustrating the main components of a portable user device of an alternative embodiment; and

Figure 5 is a schematic block diagram illustrating the main components of the remote server used with the portable user device shown in Figure 5.

OVERVIEW

Figure 1 schematically illustrates a captioning system for use in providing text captions on a number of user devices (two of which are shown and labelled 1-1 and 1-2) for a film being shown on a screen 3 within a cinema 5. The captioning system also includes a remote web server 7 which controls access by the user devices 1 to captions stored in a captions database 9. In particular, in this embodiment, the user device 1-1 is a mobile telephone which can connect to the remote web server 7 via a cellular communications base station 11, a switching centre 13 and the Internet 15 to download captions from the captions database 9. In this embodiment, the second user device 1-2 is a personal digital assistant (PDA) that does not have cellular telephone transceiver circuitry. This PDA 1-2 can, however, connect to the remote web server 7 via a computer 17 which can connect to the Internet 15. The computer 17 may be a home computer located in the user's home 19 and may typically include a docking station 21 for connecting the PDA 1-2 with the computer 17.

In this embodiment, the operation of the captioning system using the mobile telephone 1-1 is slightly different to the operation of the captioning system using the PDA 1-2. A brief description of the operation of the captioning system using these devices will now be given.

In this embodiment, the mobile telephone 1-1 operates to download the caption for the film to be viewed at the start of the film. It does this by capturing a portion of soundtrack from the beginning of the film, generated by speakers 23-1 and 23-2, which it processes to generate a signature that is characteristic of the audio segment. The mobile telephone 1-1 then transmits this signature to the remote web server 7 via the base station 11, switching station 13 and the Internet 15. The web server 7 then identifies the film that is about to begin from the signature and retrieves the appropriate caption file together with an associated synchronisation file which it transmits back to the mobile telephone 1-1 via the Internet 15, switching centre 13 and base station 11. After the caption file and the synchronisation file have been received by the mobile telephone 1-1, the connection with the base station 11 is terminated and the mobile telephone 1-1 generates and displays the appropriate captions to the user in synchronism with the film that is shown on the screen 3. In this embodiment, the synchronisation data in the synchronisation file downloaded from the remote web server 7 defines the estimated timing of subsequent audio segments within the film and the mobile telephone 1-1 synchronises the playout of the captions by processing the audio signal of the film and identifying the actual timing of those subsequent audio segments in the film.

In this embodiment, the user of the PDA 1-2 downloads the caption for the film while they are at home 19 using their personal computer 17 in advance of the film being

shown. In particular, in this embodiment, the user types in the name of the film that they are going to see into the personal computer 17 and then sends this information to the remote web 7 server via the Internet 15. In response, the web server 7 retrieves the appropriate caption file and synchronisation file for the film which it downloads to the user's personal computer 17. The personal computer 17 then stores the caption file and the synchronisation file in the PDA 1-2 via the docking station 21. In this embodiment, the subsequent operation of the PDA 1-2 to synchronise the display of the captions to the user during the film is the same as the operation of the mobile telephone 1-1 and will not, therefore, be described again.

MOBILE TELEPHONE

A brief description has been given above of the way in which the mobile telephone 1-1 retrieves and subsequently plays out the captions for a film to a user. A more detailed description will now be given of the main components of the mobile telephone 1-1 which are shown in block form in Figure 2a. As shown, the mobile telephone 1-1 includes a microphone 41 for detecting the acoustic sound signal generated by the speakers 23 in the cinema 5 and for generating a corresponding electrical audio signal. The audio signal from the microphone 41 is then filtered by a filter 43 to remove frequency components that are not of interest. The filtered audio signal is then converted into a digital signal by the analogue to digital converter (ADC) 45 and then stored in an input buffer 47. The audio signal written into the

input buffer 47 is then processed by a signature extractor 49 which processes the audio to extract a signature that is characteristic of the buffered audio. Various processing techniques can be used by the signature extractor 49 to extract this signature. For example, the signature extractor may carry out the processing described in WO 02/11123 in the name of Shazam Entertainment Limited. In this system, a window of about 15 seconds of audio is processed to identify a number of "fingerprints" along the audio string that are representative of the audio. These fingerprints together with timing information of when they occur within the audio string forms the above described signature.

As shown in Figure 2a, the signature generated by the signature extractor is then output to an output buffer 51 and then transmitted to the remote web server 7 via the antenna 53, a transmission circuit 55, a digital to analogue converter (DAC) 57 and a switch 59.

As will be described in more detail below, the remote server 7 then processes the received signature to identify the film that is playing and to retrieve the appropriate caption file and synchronisation file for the film. These are then downloaded back to the mobile telephone 1-1 and passed, via the aerial 53, reception circuit 61 and analogue to digital converter 63 to a caption memory 65. Figure 2b schematically illustrates the form of the caption file 67 downloaded from the remote web server 7. As shown, in this embodiment, the caption file 67 includes an ordered sequence of captions

(caption(1) to caption(N)) 69-1 to 69-N. The caption file 67 also includes, for each caption, formatting information 71-1 to 71-N that defines the font, colour, etc. of the text to be displayed. The caption file 67 also includes, for each caption, a time value t_1 to t_N which defines the time at which the caption should be output to the user relative to the start of the film. Finally, in this embodiment, the caption file 67 includes, for each caption 69, a duration Δt_1 to Δt_N which defines the duration that the caption should be displayed to the user.

Figure 2c schematically represents the data within the synchronisation file 73 which is used in this embodiment by the mobile telephone 1-1 to synchronise the display of the captions with the film. As shown, the synchronisation file 73 includes a number of signatures 75-1 to 75-M each having an associated time value t_1^S to t_M^S identifying the time at which the signature should occur within the audio of the film (again calculated from the beginning of the film).

In this embodiment, the synchronisation file 73 is passed to a control unit 81 which controls the operation of the signature extracting unit 49 and a sliding correlator 83. The control unit 81 also controls the position of the switch 59 so that after the caption and synchronisation files have been downloaded into the mobile telephone 1-1, and the mobile telephone 1-1 is trying to synchronise the output of the captions with the film, the signature stream generated by the signature extractor 49 is passed

to the sliding correlator 83 via the output buffer 51 and the switch 59.

Initially, before the captions are output to the user, the mobile telephone 1-1 must synchronise with the film that is playing. This is achieved by operating the signature extractor 49 and the sliding correlator 83 in an acquisition mode, during which the signature extractor extracts signatures from the audio received at the microphone 41 which are then compared with the signatures 75 in the synchronisation file 73, until it identifies a match between the received audio from the film and the signatures 75 in the synchronisation file 73. This match identifies the current position within the film, which is used to identify the initial caption to be displayed to the user. At this point, the mobile telephone 1-1 enters a tracking mode during which the signature extractor 49 only extracts signatures for the audio during predetermined time slots (or windows) within the film corresponding to when the mobile telephone 1-1 expects to detect the next signature in the audio track of the film. This is illustrated in Figure 2d which shows a time line (representing the time line for the film) together with the timings t_1^S to t_M^S corresponding to when the mobile telephone 1-1 expects the signatures to occur within the audio track of the film. Figure 2d also shows a small time slot or window w_1 to w_M around each of these time points, during which the signature extractor 49 processes the audio signal to generate a signature stream which it outputs to the output buffer 51.

The generation of the signature stream is illustrated in Figure 2e which shows a portion 77 of the audio track corresponding to one of the time windows w_j and the stream 79 of signatures generated by the signature extractor 49. In this embodiment, three signatures (signature (i), signature (i+1) and signature (i+2)) are generated for each processing window w. This is for illustration purposes only. In practice, many more or less signatures may be generated for each processing window w. Further, whilst in this embodiment the signatures are generated from non-overlapping subwindows of the processing window w, the signatures may also be generated from overlapping subwindows. The way in which this would be achieved will be well known to those skilled in the art and will not be described in any further detail.

In this embodiment, between adjacent processing windows w, the control unit 51 controls the signature extractor 49 so that it does not process the received audio. In this way, the processing performed by the signature extractor 49 can be kept to a minimum.

During this tracking mode of operation, the sliding correlator 83 is operable to correlate the generated signature stream in output buffer 51 with the next signature 75 in the synchronisation file 73. This correlation generates a correlation plot such as that shown in Figure 2f for the window of audio being processed. As shown in Figure 2d, in this embodiment, the windows w_j are defined so that the expected timing of

the signature is in the middle of the window. This means that the mobile telephone 1-1 expects the peak output from the sliding correlator 83 to correspond to the middle of the processing window w. If the peak occurs earlier or later in the window then the caption output timing of the mobile telephone 1-1 must be adjusted to keep it in synchronism with the film. This is illustrated in Figure 2f which shows the expected time of the signature t_s appearing in the middle of the window and the correlation peak occurring δt seconds before the expected time. This means that the mobile telephone 1-1 is slightly behind the film and the output timing of the subsequent captions must be brought forward to catch up with the film. This is achieved by passing the δt value from the correlator 83 into a timing controller 85 which generates the timing signal for controlling the time at which the captions are played out to the user. As shown, the timing controller receives its timing reference from the mobile telephone clock 87. The generated timing signal is then passed to a caption display engine 89 which uses the timing signal to index the caption file 67 in order to retrieve the next caption 69 for display together with the associated duration information Δt and formatting information 71 which it then processes and outputs for display on the mobile telephone display 91 via a frame buffer 93. The details of how the caption 69 is generated and formatted are well known to those skilled in the art and will not be described in any further detail.

Figure 2g illustrates the form of an example caption

which is output on the display 91. Figure 2g also shows in the right hand side 95 of the display 91 a number of user options that the user can activate by pressing appropriate function keys on the keypad 97 of the mobile telephone 1-1. These include a language option 99 which allows the user to change the language of the caption 69 that is displayed. This is possible, provided the caption file 67 includes captions 69 in different languages. As the skilled man will appreciate, this does not involve any significant processing on the part of the mobile telephone 1-1, since all that is being changed is the text of the caption 69 that is to be displayed at the relevant timings. It is therefore possible to personalise the captions for different users watching the same film. The options also include an exit option 101 for allowing the user to exit the captioning application being run on the mobile telephone 1-1.

PERSONAL DIGITAL ASSISTANT

As mentioned above, the PDA 1-2 operates in a similar way to the mobile telephone 1-1 except it does not include the mobile telephone transceiver circuitry for connecting directly to the web server 7. The main components of the PDA 1-2 are similar to those of the mobile telephone 1-1 described above and will not, therefore, be described again.

REMOTE WEB SERVER

Figure 3 is a schematic block diagram illustrating the main components of the web server 7 used in this embodiment and showing the captions database 9. As

shown, the web server 7 receives input from the Internet 15 which is either passed to a sliding correlator 121 or to a database reader 123, depending on whether or not the input is from the mobile telephone 1-1 or from the PDA 1-2. In particular, the signature from the mobile telephone 1-1 is input to the sliding correlator 121 where it is compared with signature streams of all films known to the system, which are stored in the signature stream database 125. The results of these correlations are then compared to identify the film that the user is about to watch. This film ID is then passed to the database reader 123. In response to receiving a film ID either from the sliding correlator 121 or directly from a user device (such as the PC 17 or PDA 1-2), the database reader 123 reads the appropriate caption file 67 and synchronisation file 73 from the captions database 9 and outputs them to a download unit 127. The download unit 127 then downloads the retrieved caption file 67 and synchronisation file 73 to the requesting user device 1 via the Internet 15.

As those skilled in the art will appreciate, a captioning system has been described above for providing text captions for a film for display to a user. The system does not require any modifications to the cinema or playout system, but only the provision of a suitably adapted mobile telephone 1-1 or PDA device 1-2 or the like. In this regard, it is not essential to add any additional hardware to the mobile telephone or the PDA, since all of the functionality enclosed in the dashed box 94 can be performed by an appropriate software

application run within the mobile telephone 1-1 or PDA 1-2. In this case, the appropriate software application may be loaded at the appropriate time, e.g. when the user enters the cinema and in the case of the mobile telephone 1-1, is arranged to cancel the ringer on the telephone so that incoming calls do not disturb others in the audience. The above captioning system can therefore be used for any film at any time. Further, since different captions can be downloaded for a film, the system allows for content variation within a single screening. This facilitates, for example, the provision of captions in multiple languages.

Modifications and Alternative Embodiments

In the above embodiment, a captioning system was described for providing text captions on a display of a portable user device for allowing users with hearing disabilities to understand a film being watched. As discussed in the introduction of this application, the above captioning system can be modified to operate with audio captions (e.g. audio descriptions of the film being displayed for people with impaired eyesight). This may be done simply by replacing the text captions 69 in the caption file 67 that is downloaded from the remote server 7 with appropriate audio files (such as the standard .WAV or MP3 audio files) which can then be played out to the user via an appropriate headphone or earpiece. The synchronisation of the playout of the audio files could be the same as for the synchronisation of the playout of the text captions. Alternatively synchronisation can be achieved in other ways. Figure 4 is a block diagram

illustrating the main components of a mobile telephone that can be used in such an audio captioning system. In Figure 4, the same reference numerals have been used for the same components shown in Figure 2a and these components will not be described again.

In this embodiment, the mobile telephone 1-1' does not include the signature extractor 49. Instead, as illustrated in Figure 5, the signature extractor 163 is provided in the remote web server 7'. In operation, the mobile telephone 1-1' captures part of the audio played out at the beginning of the film and transmits this audio through to the remote web server 7'. This audio is then buffered in the input buffer 161 and then processed by the signature extractor 163 to extract a signature representative of the audio. This signature is then passed to a correlation table 165 which performs a similar function to the sliding correlator 121 and signature stream database 125 described in the first embodiment, to identify the ID for the film currently being played. In particular, in this embodiment, all of the possible correlations that may have been performed by the sliding correlator 121 and the signature stream database 125 are carried out in advance and the results are stored in the correlation table 165. In this way, the signature output by the signature extractor 163 is used to index this correlation table to generate correlation results for the different films known to the captioning system. These correlation results are then processed to identify the most likely film corresponding to the received audio stream. In this embodiment, the

captions database 9 only includes the caption files 67 for the different films, without any synchronisation 73 files. In response to receiving the film ID either from the correlation table 165 or from a user direct from a user device (not shown), the database reader 123 retrieves the appropriate caption file 67 which it downloads to the user device 1 via the download unit 127.

Returning to Figure 4, in this embodiment, since the mobile telephone 1-1' does not include the signature extractor 49, synchronisation is achieved in an alternative manner. In particular, in this embodiment, synchronisation codes are embedded within the audio track of the film. Therefore, after the caption file 67 has been stored in the caption memory 65, the control circuit 81 controls the position of the switch 143 so that the audio signal input into the input buffer 47 is passed to a data extractor 145 which is arranged to extract the synchronisation data that is embedded in the audio track. The extracted synchronisation data is then passed to the timing controller 85 which controls the timing at which the individual audio captions are played out by the caption player 147 via the digital-to-analogue converter 149, amplifier 151 and the headset 153.

As those skilled in the art will appreciate, various techniques can be used to embed the synchronisation data within the audio track. The applicant's earlier International applications WO 98/32248, WO 01/10065, PCT/GB01/05300 and PCT/GB01/05306 describe techniques for embedding data within acoustic signals and appropriate

data extractors for subsequently extracting the embedded data. The contents of these earlier International applications are incorporated herein by reference.

In the above audio captioning embodiment, synchronisation was achieved by embedding synchronisation codes within the audio and detecting these in the mobile telephone. As those skilled in the art will appreciate, a similar technique may be used in the first embodiment. However, embedding audio codes within the soundtrack of the film is not preferred, since it involves modifying in some way the audio track of the film. Depending on the data rates involved, this data may be audible to some viewers which may detract from their enjoyment of the film. The first embodiment is therefore preferred since it does not involve any modification to the film or to the cinema infrastructure.

In embodiments where the synchronisation data is embedded within the audio, the synchronisation codes used can either be the same code repeated whenever synchronisation is required or it can be a unique code at each synchronisation point. The advantage of having a unique code at each synchronisation point is that a user who enters the film late or who requires the captions only at certain points (for example a user who only rarely requires the caption) can start captioning at any point during the film.

In the embodiment described above with reference to Figures 4 and 5, the signature extraction operation was

performed in the remote web server rather than in the mobile telephone. As those skilled in the art will appreciate, this modification can also be made to the first embodiment described above, without the other modifications described with reference to Figures 4 and 5.

In the first embodiment described above, during the tracking mode of operation, the signature extractor only processed the audio track during predetermined windows in the film. As those skilled in the art will appreciate, this is not essential. The signature extractor could operate continuously. However, such an embodiment is not preferred since it increases the processing that the mobile telephone has to perform which is likely to increase the power consumption of the mobile telephone.

In the above embodiments, the mobile telephone or PDA monitored the audio track of the film for synchronisation purposes. As those skilled in the art will appreciate, the mobile telephone or PDA device may be configured to monitor the video being displayed on the film screen. However, this is currently not preferred because it would require an image pickup device (such as a camera) to be incorporated into the mobile telephone or PDA and relatively sophisticated image processing hardware and software to be able to detect the synchronisation points or codes in the video. Further, it is not essential to detect synchronisation codes or synchronisation points from the film itself. Another electromagnetic or

pressure wave signal may be transmitted in synchronism with the film to provide the synchronisation points or synchronisation codes. In this case, the user device would have to include an appropriate electromagnetic or pressure wave receiver. However, this embodiment is not preferred since it requires modification to the existing cinema infrastructure and it requires the generation of the separate synchronisation signal which is itself synchronised to the film.

In the above embodiments, the captions and where appropriate the synchronisation data, were downloaded to a user device from a remote server. As those skilled in the art will appreciate, the use of such a remote server is not essential. The caption data and the synchronisation data may be pre-stored in memory cards or smart cards and distributed or sold at the cinema. In this case, the user device would preferably have an appropriate slot for receiving the memory card or smart-card and an appropriate reader for accessing the caption data and, if provided, the synchronisation data. The manufacture of the cards would include the steps of providing the memory card or smart-card and using an appropriate card writer to write the captions and synchronisation data into the memory card or into a memory on the smart-card. Alternatively still, the user may already have a smart-card or memory card associated with their user device which they simply insert into a kiosk at the cinema where the captions and, if applicable, the synchronisation data are written into a memory on the card.

As a further alternative, the captions and synchronisation data may be transmitted to the user device from a transmitter within the cinema. This transmission may be over an electromagnetic or a pressure wave link.

In the first embodiment described above, the mobile telephone had an acquisition mode and a subsequent tracking mode for controlling the playout of the captions. In an alternative embodiment, the acquisition mode may be dispensed with, provided that the remote server can identify the current timing from the signature received from the mobile telephone. This may be possible in some instances. However, if the introduction of the film is repetitive then it may not be possible for the web server to be able to provide an initial synchronisation.

In the first embodiment described above, the user devices downloaded the captions and synchronisation data from a remote web server via the internet. As those skilled in the art will appreciate, it is not essential to download the files over the internet. The files may be downloaded over any wide area or local area network. The ability to download the caption files from a wide area network is preferred since centralised databases of captions may be provided for distribution over a wider geographic area.

In the first embodiment described above, the user downloaded captions and synchronisation data from a

remote web server. Although not described, for security purposes, the caption file and the synchronisation file are preferably encoded or encrypted in some way to guard against fraudulent use of the captions. Additionally, the caption system may be arranged so that it can only operate in cinemas or at venues that are licensed under the captioning system. In this case, an appropriate activation code may be provided at the venue in order to "unlock" the captioning system on the user device. This activation may be provided in human readable form so that the user has to key in the code into the user device. Alternatively, the venue may be arranged to transmit the code (possibly embedded in the film) to an appropriate receiver in the user device. In either case, the captioning system software in the user device would have an inhibitor that would inhibit the outputting of the captions until it received the activation code. Further, where encryption is used, the activation code may be used as part of the key for decrypting the captions.

The above embodiments have described text captioning systems and audio captioning systems for use in a cinema. As those skilled in the art will appreciate, these captioning systems may be used for providing captions for any radio, video or multi-media presentation. They can also be used in the theatre or opera or within the user's home.

Various captioning systems have been described above which provide text or audio captions for an audio or a video presentation. The captions may include extra

commentary about the audio or video presentation, such as director's comments, explanation of complex plots, the names of actors in the film or third party comments. The captions may also include adverts for other products or presentations. In addition, the audio captioning system may be used not only to provide audio descriptions of what is happening in the film, but also to provide a translation of the audio track for the film. In this way, each listener in the film can listen to the film in their preferred language. The caption system can also be used to provide karaoke captions for use with standard audio tracks. In this case, the user would download the lyrics and the synchronisation information which define the timing at which the lyrics should be displayed and highlighted to the user.

In addition to the above, the captioning system described above may be provided to control the display of video captions. For example, such video captions can be used to provide sign language (either real images or computer generated images) for the audio in the presentation being given.

In the above embodiments, the captions for the presentation to be made were downloaded in advance for playout. In an alternative embodiment, the captions may be downloaded from the remote server by the user device when they are needed. For example, the user device may download the next caption when it receives the next synchronisation code for the next caption.

In the caption system described above, a user downloads or receives the captions and the synchronisation information either from a web server or locally at the venue at which the audio or visual presentation is to be made. As those skilled in the art will appreciate, for applications where the user has to pay to download or playout the captions, a transaction system is preferably provided to facilitate the collection of the monies due. In embodiments where the captions are downloaded from a web server, this transaction system preferably forms part of or is associated with the web server providing the captions. In this case, the user can provide electronic payment or payment through credit card or the like at the time that they download the captions. This is preferred, since it is easier to link the payment being made with the captions and synchronisation information downloaded.

In the first embodiment described above, the ID for the film was automatically determined from an audio signature transmitted from the user's mobile telephone. Alternatively, instead of transmitting the audio signature, the user can input the film ID directly into the telephone for transmission to the remote server. In this case, the correlation search of the signature database is not essential.

In the first embodiment described above, the user device processed the received audio to extract a signature characteristic of the film that they are about to watch. The processing that is preferred is the processing described in the Shazam Entertainment Ltd patent

mentioned above. However, as those skilled in the art will appreciate, other types of encoding may be performed. The main purpose of the signature extractor unit in the mobile telephone is to compress the audio to generate data that is still representative of the audio from which the remote server can identify the film about to be watched. Various other compression schemes may be used. For example, a GSM codec together with other audio compression algorithms may be used.

In the above embodiments in which text captions are provided, they were displayed to the user on a display of a portable user device. Whilst this offers the simplest deployment of the captioning system, other options are available. For example, the user may be provided with an active or passive type head-up-display through which the user can watch the film and on which the captions are displayed (active) or are projected (passive) to overlay onto the film being watched. This has the advantage that the user does not have to watch two separate displays. A passive type of head-up-display can be provided, for example, by providing the user with a pair of glasses having a beam splitter (e.g. a 45° prism) on which the user can see the cinema screen and the screen of their user device (e.g. phone or PDA) sitting on their lap. Alternatively, instead of using a head-up-display, a separate transparent screen may be erected in front of the user's seat and onto which the captions are projected by the user device or a seat-mounted projector.

In the first embodiment described above, the caption file included a time ordered sequence of captions together with associated formatting information and timing information. As those skilled in the art will appreciate, it is not essential to arrange the captions in such of time sequential order. However, arranging them in this way reduces the processing involved in identifying the next caption to display. Further, it is not essential to have formatting information in addition to the caption. The minimum information required is the caption information. Further, it is not essential that this be provided in a file as each of the individual captions for the presentation may be downloaded separately. However, the above described format for the caption file is preferred since it is simple and can easily be created using, for example, a spreadsheet. This simplicity also provides the potential to create a variety of different caption content.

In embodiments where the user's mobile telephone is used to provide the captioning, the captioning system can be made interactive whereby the user can interact with the remote server, for example interacting with adverts or questionaries before the film starts. This interaction can be implemented using, for example, a web browser on the user device that receives URLs and links to other information on websites.

In the first embodiment described above, text captions were provided for the audio in the film to be watched. These captions may include full captions, subtitles for

the dialogue only or subtitles at key parts of the plot. Similar variation may be applied for audio captions.

CLAIMS:

1. A captioning system for providing captions for a presentation to a user, the captioning system comprising:

a caption store operable to store one or more sets of captions each set being associated with one or more presentations and each set comprising a plurality of captions for playout at different timings during the associated presentation; and

a user device having:

i) a memory operable to receive and store at least one set of captions for a presentation to be made to an associated user, from said caption store;

ii) a receiver operable to receive synchronisation information defining the timing during the presentation at which each caption in the received set of captions is to be output to the user;

iii) a caption output circuit operable to output to the associated user, the captions in the received set of captions; and

iv) a timing controller responsive to said received synchronisation information and operable to control said caption output circuit so that said captions are output to said user at the timings defined by said synchronisation information.

2. A system according to claim 1, wherein said captions include text.

3. A system according to claim 2, wherein said captions include text for any dialogue in the presentation.

4. A system according to claim 2 or 3, wherein said caption output circuit is operable to output said captions to a display device associated with the user device for display to the user.

5. A system according to claim 4, wherein said captions include formatting information for controlling the format of the text displayed on said display.

6. A system according to claim 4 or 5, wherein each caption includes duration information defining the duration that the caption should be displayed to the user.

7. A system according to any of claims 4 to 6, wherein said caption includes timing information defining the time at which the caption should be displayed to the user during the presentation.

8. A system according to any preceding claim wherein said captions include audio data and wherein said caption output circuit is operable to output said audio data to an electro-acoustic device for converting the audio data into corresponding acoustic signals.

9. A system according to any preceding claim, wherein said presentation includes audio.

10. A system according to any preceding claim, wherein said presentation includes video.

11. A system according to any preceding claim, wherein said presentation is a film.

12. A system according to any preceding claim, wherein said caption store is formed in a memory card which is insertable into said user device and wherein said user device includes a reader for reading captions from said memory card when inserted therein.

13. A system according to any of claims 1 to 11, wherein said caption store is provided in a computer system and wherein said user device includes means for communicating with said computer system.

14. A system according to claim 13, wherein said computer system is remote from said user device and wherein said user device has an associated communication module for communicating with said remote computer system.

15. A system according to claim 14, wherein said user device includes a housing and wherein said communication module is provided within said housing.

16. A system according to claim 14 or 15, wherein said communication module is operable to communicate with said remote computer system using a wireless communication link.

17. A system according to claim 16, wherein said user device comprises a mobile telephone.

18. A system according to any proceeding claim, wherein said user device comprises a portable computing device such as a personal digital assistant.

19. A system according to any preceding claim, wherein said synchronisation information defines expected time points for one or more predetermined portions of the presentation.

20. A system according to claim 19, wherein said user device comprises a monitoring circuit operable to monitor said presentation to identify the actual time points of said one or more predetermined portions and wherein said timing controller is responsive to the difference between the actual timings and the expected timings to control the outputting of the captions by said caption output circuit.

21. A system according to claim 20, wherein said predetermined portions of said presentation correspond to portions of audio of the presentation and wherein said monitoring circuit includes a microphone for sensing the audio of the presentation and a comparator for comparing the received audio with the expected portions of the audio defined by said synchronisation information.

22. A system according to claim 20 or 21, wherein said user device has an acquisition mode of operation in which an output of said monitoring circuit is compared with said predetermined points defined by said synchronisation information to identify a current position within said

presentation and a tracking mode of operation in which the output of said monitoring circuit is compared with a current predetermined portion defined by said synchronisation information.

23. A system according to claim 22, wherein during said tracking mode of operation, said monitoring circuit is operable to monitor said presentation during a predetermined time window around the expected time point defined by said synchronisation information for the current predetermined portion.

24. A system according to any preceding claim, wherein said receiver in said user device is operable to receive said synchronisation information from said caption store.

25. A system according to any of claims 1 to 23, wherein said synchronisation information is embedded within said presentation and wherein said user device includes a monitoring circuit operable to monitor the presentation and to extract said synchronisation information therefrom.

26. A system according to claim 25, wherein said synchronisation information is embedded within the audio of said presentation.

27. A system according to claim 25 or 26, wherein said synchronisation information comprises synchronisation codes occurring at different timings during the presentation.

28. A system according to claim 27, wherein each synchronisation code is unique to uniquely define the position in the presentation.

29. A system according to any preceding claim, wherein said caption store includes a plurality of sets of captions for a plurality of different presentations.

30. A system according to claim 29, wherein said user device is operable to capture a portion of said presentation and is operable to transmit the captured portion to said caption store and when said caption store is operable to use said captured portion of the presentation to identify the presentation being made and to transmit the associated set of captions for the identified presentation to said user device.

31. A system according to claim 30, wherein said user device is operable to process the captured portion of the presentation to extract data characteristic of the captured portion and is operable to transmit said characteristic data to said caption store, and wherein said caption store is operable to use said characteristic data to identify the presentation being made and to transmit the associated set of captions for the identified presentation to the user device.

32. A system according to any preceding claim, wherein said presentation is given at a venue, wherein said venue is operable to provide an activation code, wherein said user device is operable to receive said activation code

and further comprises an inhibitor for inhibiting the operation of said caption output circuit unless said user device has received said activation code.

33. A user device for use in a captioning system, the user device comprising:

i) a memory operable to receive and store at least one set of captions for a presentation to be made to an associated user, from said caption store;

ii) a receiver operable to receive synchronisation information defining the timing during the presentation at which each caption in the received set of captions is to be output to the user;

iii) a caption output circuit operable to output to the associated user, the captions in the received set of captions; and

iv) a timing controller responsive to said received synchronisation information and operable to control said caption output circuit so that said captions are output to said user at the timings defined by said synchronisation information.

34. A computer system for use in a captioning system, the computer system comprising a caption store operable to store one or more sets of captions each set being associated with one or more presentations and each set comprising a plurality of captions which playout at different timings during the associated presentation and each caption having associated synchronisation information defining the timing during the presentation in which each caption in the received set of captions is

to be output to the user;

a receiver operable to receive a request for a set of captions from a user device; and

an output circuit operable to output the requested set of captions and the synchronisation information to the user device.

35. A method of manufacturing a computer readable medium storing caption data and synchronisation data for use in a captioning system, the method comprising:

providing a computer readable medium;

providing a set of captions that is associated with a presentation which comprises a plurality of captions for playout at different timings during the associated presentation;

providing synchronisation information defining the timing during the presentation at which each caption in the set of captions is to be output to a user;

receiving a computer readable medium;

recording computer readable data defining said set of captions and said synchronisation information on said computer readable medium; and

outputting the computer readable medium having the recorded caption and synchronisation data thereon.

36. A computer readable medium storing computer executable instructions for causing a general purpose computing device to operate as the user device in any of claims 1 to 33.

37. A method of providing captions for presentation to

a user, the method comprising:

storing, at a caption store, one or more sets of captions each being associated with one or more presentations and each comprising a plurality of captions for playout at different timings during the associated presentation; and

at a user device:

receiving and storing at least one set of captions for a presentation to be made to an associated user from said caption store;

receiving synchronisation information defining the timing during the presentation at which each caption in the received set of captions is to be output to the user;

outputting the captions in the received set of captions to the associated user; and

in response to the received synchronisation information controlling the outputting step so that said captions are output to the user at the timings defined by the synchronisation information.

38. A captioning system for providing captions for a presentation to a user, the captioning system comprising:

a caption store operable to store one or more sets of captions each set being associated with one or more presentations and each set comprising one or more captions for playout during the associated presentation; and

a user device having:

i) a memory operable to receive and store at least one set of captions for a presentation to be made to an associated user, from said caption store;

ii) a receiver operable to receive synchronisation information defining the timing during the presentation at which the or each caption in the received set of captions is to be output to the user; and

iii) a caption output circuit operable to output to the associated user, the or each caption in the received set of captions; and

iv) a timing controller responsive to said received synchronisation information and operable to control said caption output circuit so that the or each caption is output to said user at the timing defined by said synchronisation information.

39. A captioning system for providing captions for a presentation to a user, the captioning system comprising:

a caption store operable to store one or more sets of captions each set being associated with one or more presentations and each set comprising a plurality of captions for playout at different timings during the associated presentation; and

a user device having:

i) a memory operable to receive and store at least one set of captions for a presentation to be made to an associated user, from said caption store;

ii) a receiver operable to receive synchronisation information defining the timing during the presentation at which each caption in the received set of captions is to be output to the user; and

iii) a caption output circuit operable to output to the associated user, the captions in the received set of captions at the timings defined by said synchronisation

information.

40. A captioning system for providing captions for a presentation to a user, the captioning system comprising:

means for storing one or more sets of captions each set being associated with one or more presentations and each set comprising a plurality of captions for playout at different timings during the associated presentation; and

a user device having:

i) means for receiving captions from said captions store;

ii) means for receiving synchronisation information defining the timing during the presentation at which each caption is to be output to the user;

iii) means for outputting the captions to a user associated with the user device; and

iv) means responsive to the synchronisation information for controlling said output means, so that said captions are output to said user at the timings defined by said synchronisation information.

41. A computer readable medium storing caption data and synchronisation data for a presentation, the caption data defining a set of captions for the presentation and comprising a plurality of captions for playout at different timings during the presentation; and synchronisation data defining the timing during the presentation at which each caption in the received set of captions is to be output to a user.

1/7

FIG. 1

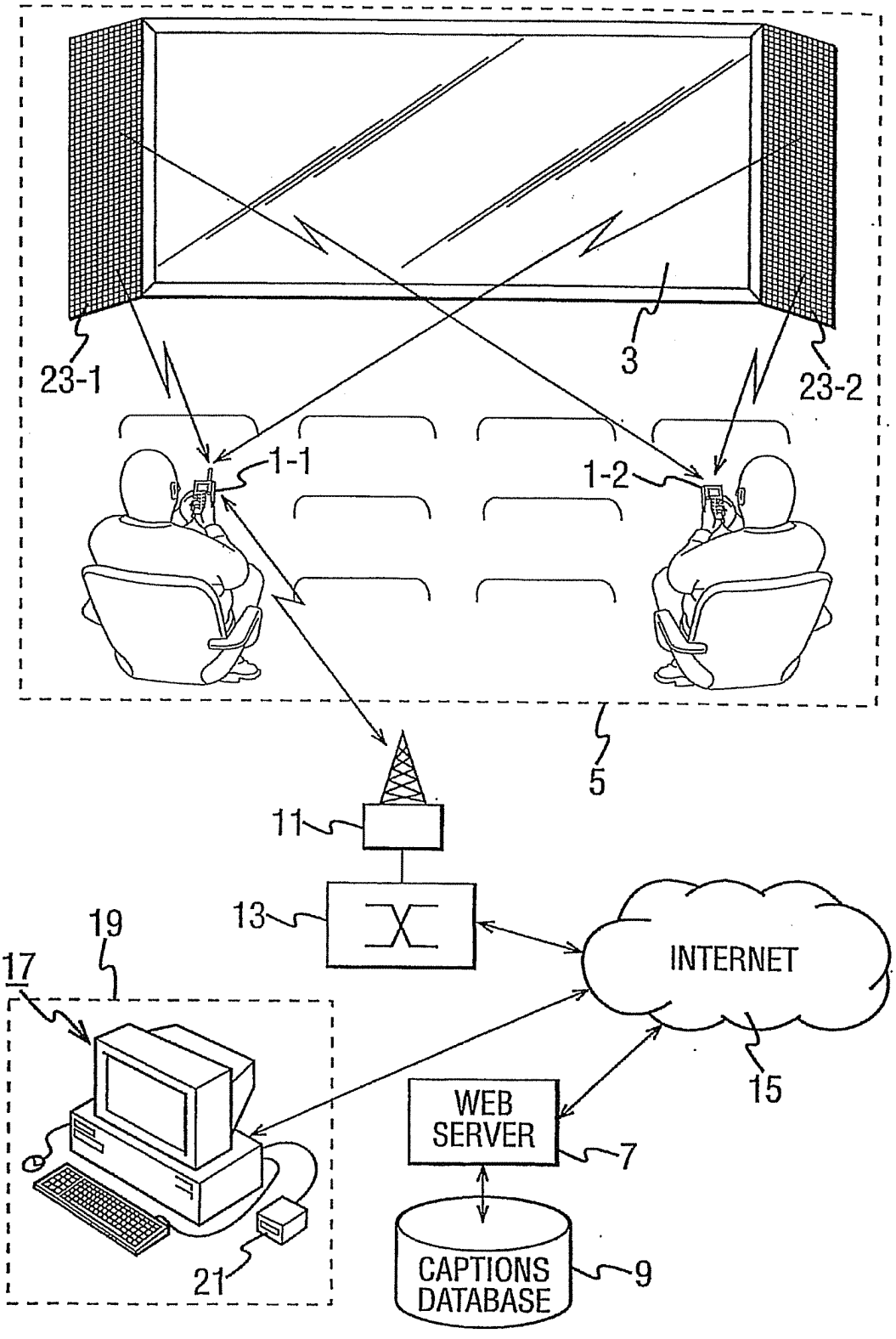

23-1
3
23-2
1-1
1-2
5
11
13
19
17
INTERNET
15
WEB SERVER
7
CAPTIONS DATABASE
9
21

FIG. 2a

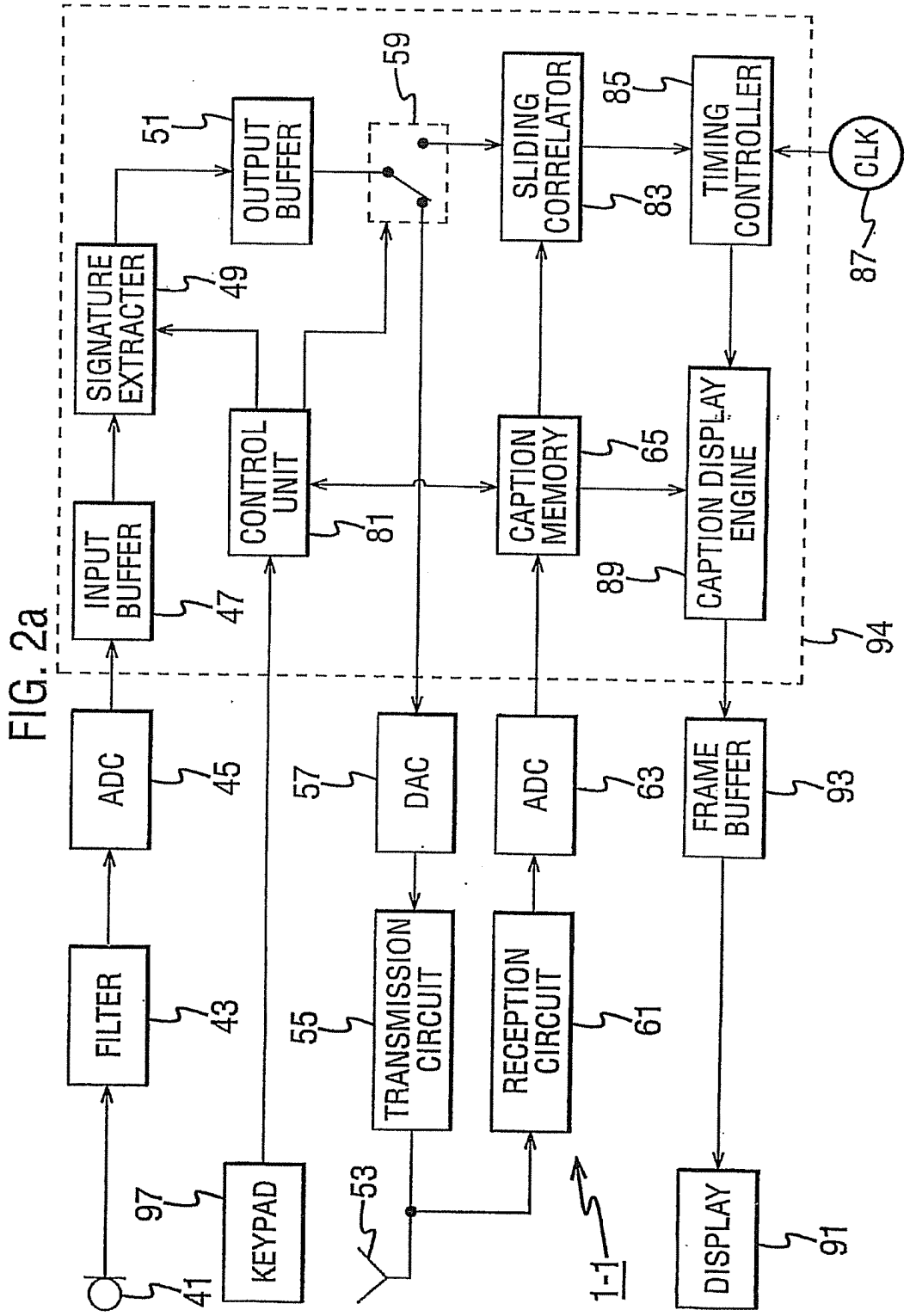

41
97
KEYPAD
53
1-1
FILTER
43
ADC
45
INPUT BUFFER
47
SIGNATURE EXTRACTER
49
51
OUTPUT BUFFER
59
CONTROL UNIT
81
55
TRANSMISSION CIRCUIT
57
DAC
RECEPTION CIRCUIT
61
ADC
63
CAPTION MEMORY
65
SLIDING CORRELATOR
83
85
TIMING CONTROLLER
87
CLK
89
CAPTION DISPLAY ENGINE
94
FRAME BUFFER
93
DISPLAY
91

FIG. 2b

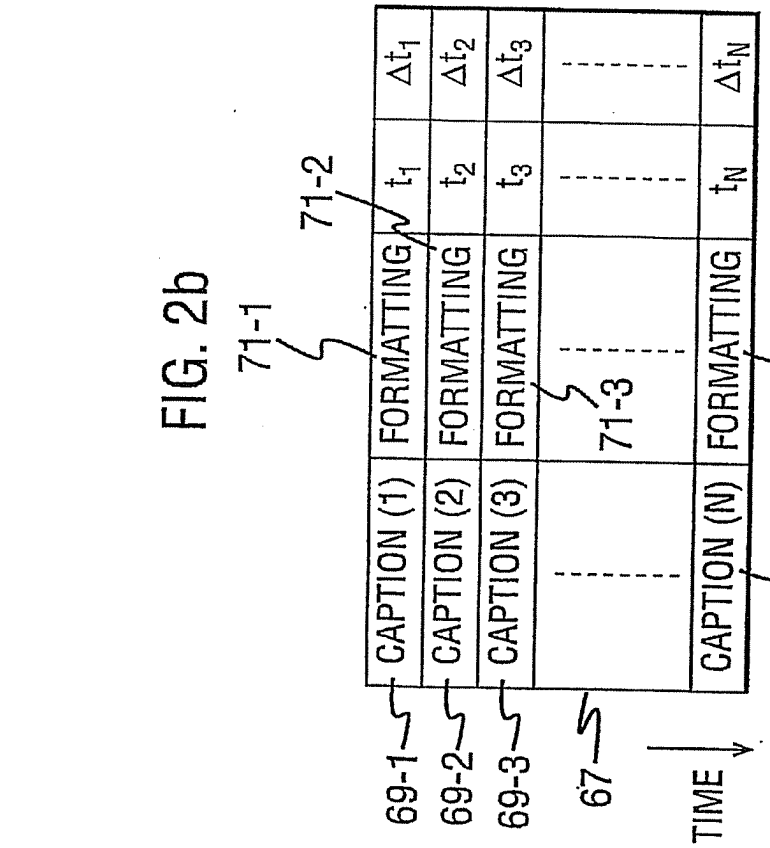

71-1
71-2
71-3
71-N
CAPTION (1)
CAPTION (2)
CAPTION (3)
CAPTION (N)
FORMATTING
FORMATTING
FORMATTING
FORMATTING
t_1
t_2
t_3
t_N
Δt_1
Δt_2
Δt_3
Δt_N
69-1
69-2
69-3
69-N
67
TIME

FIG. 2c

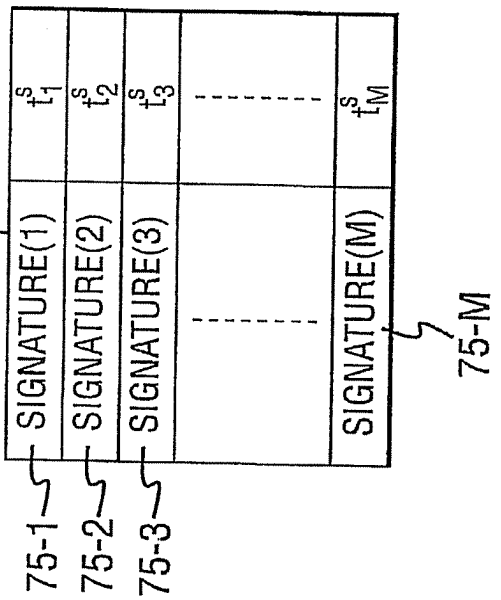

73
SIGNATURE(1)
SIGNATURE(2)
SIGNATURE(3)
SIGNATURE(M)
t^S_1
t^S_2
t^S_3
t^S_M
75-1
75-2
75-3
75-M
TIME

FIG. 2d

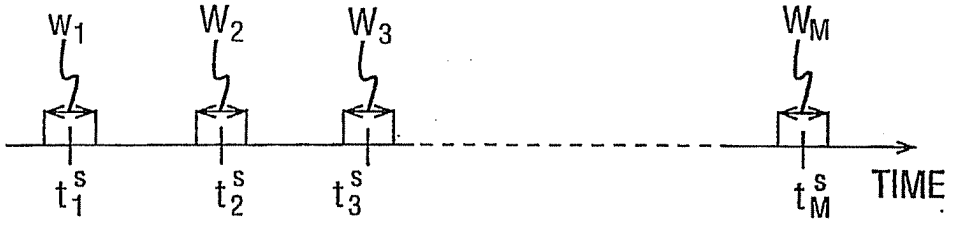


FIG. 2e

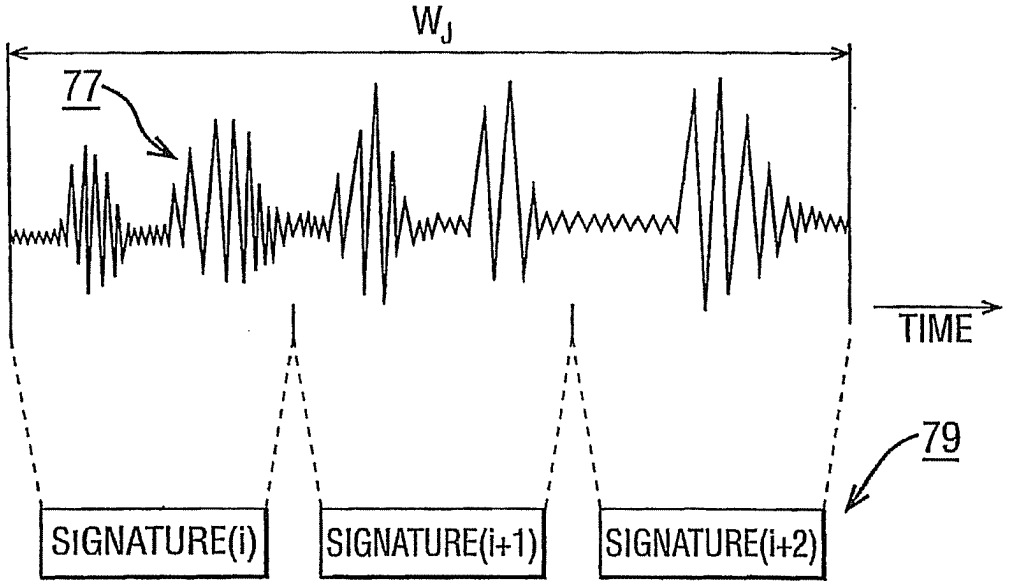


FIG. 2f

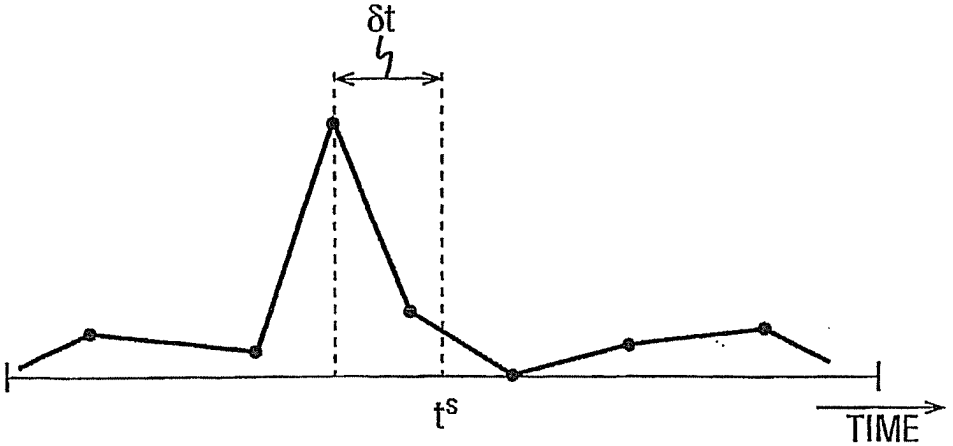


FIG. 2g

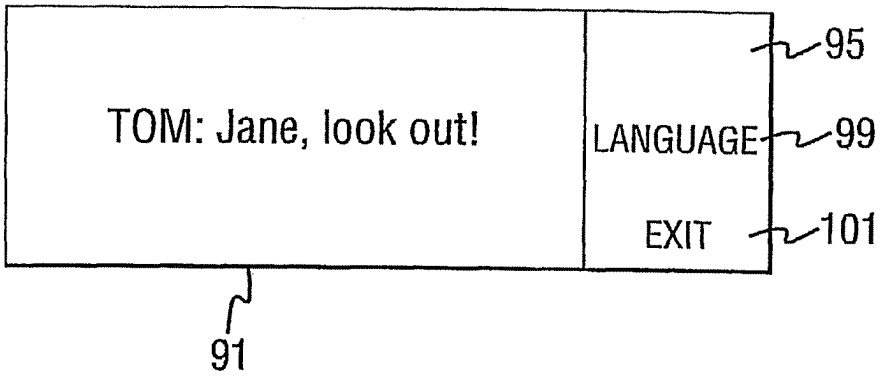

TOM: Jane, look out!
LANGUAGE
EXIT
95
99
101
91

FIG. 3

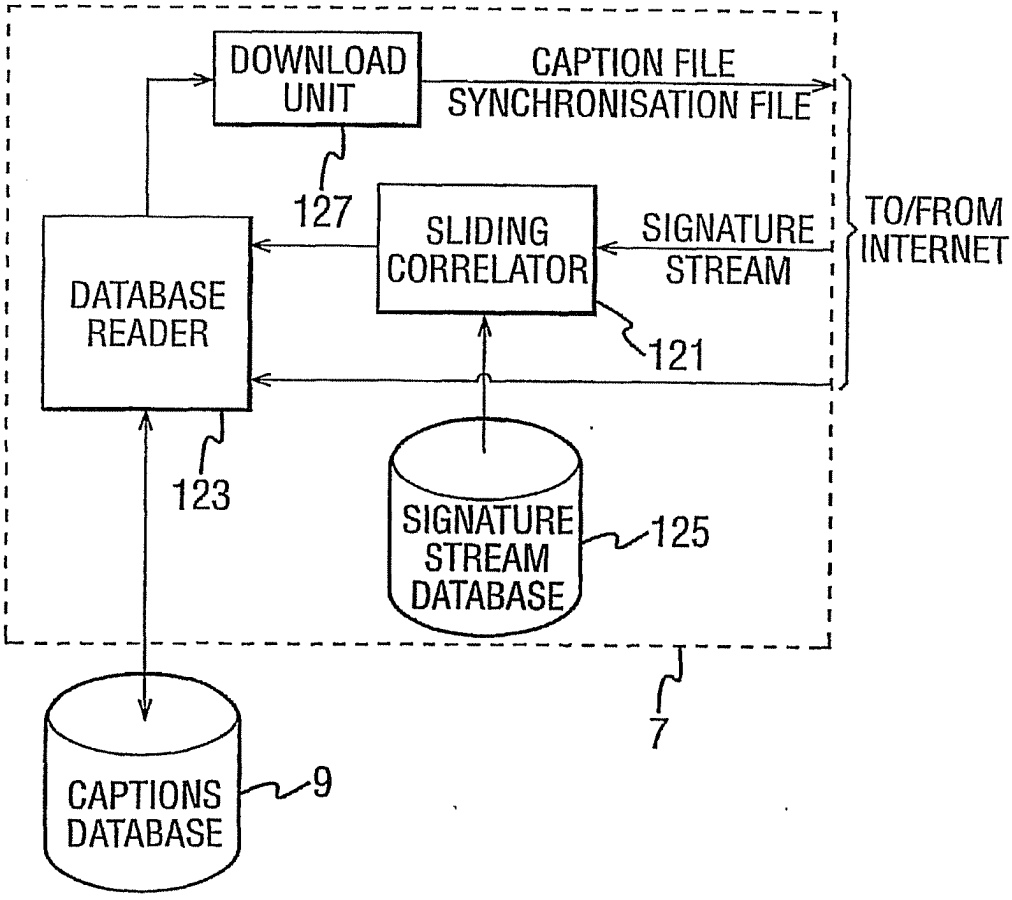

DOWNLOAD UNIT
CAPTION FILE
SYNCHRONISATION FILE
127
SLIDING CORRELATOR
SIGNATURE STREAM
TO/FROM INTERNET
DATABASE READER
121
123
SIGNATURE STREAM DATABASE
125
7
CAPTIONS DATABASE
9

FIG. 4

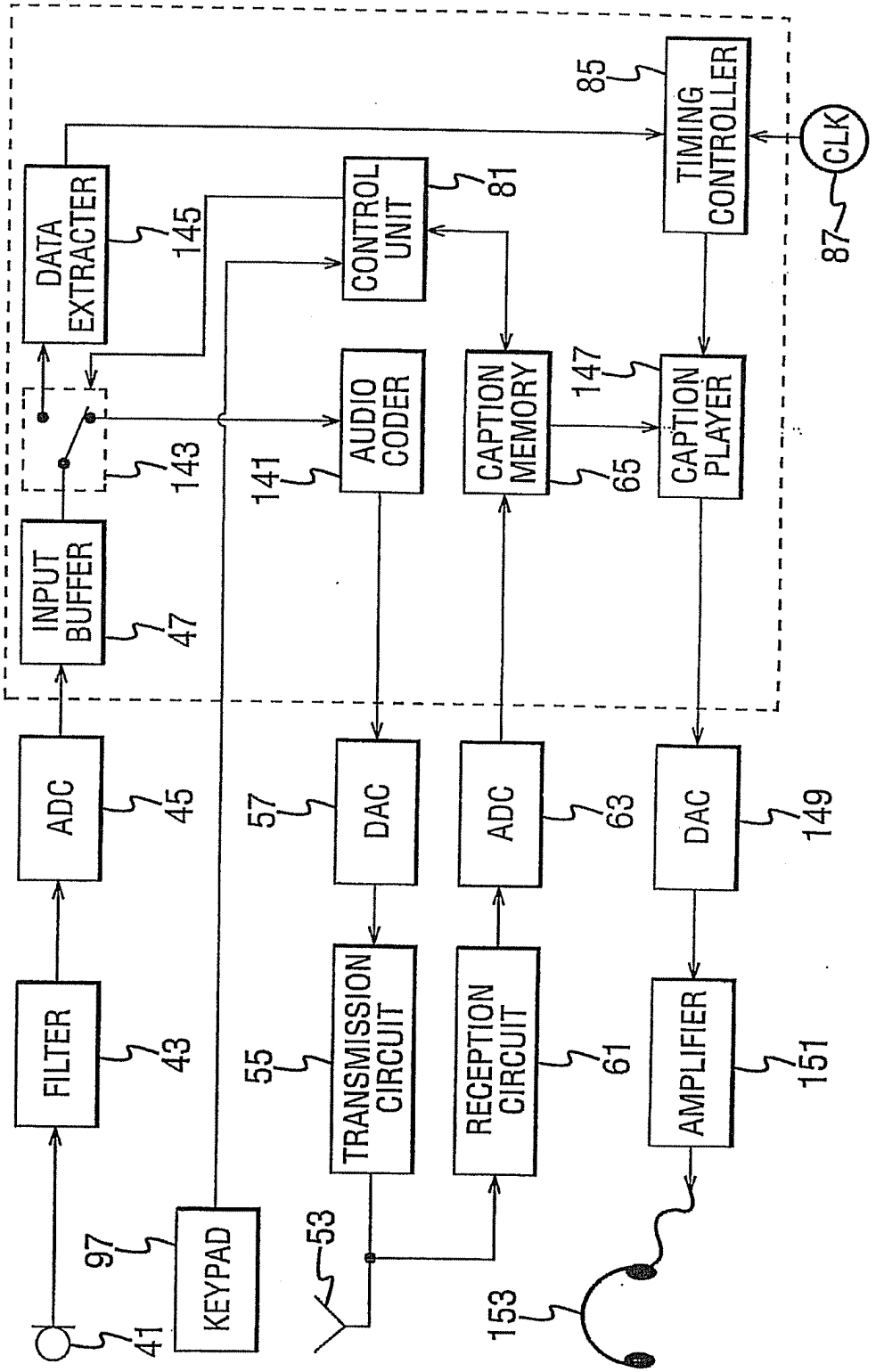
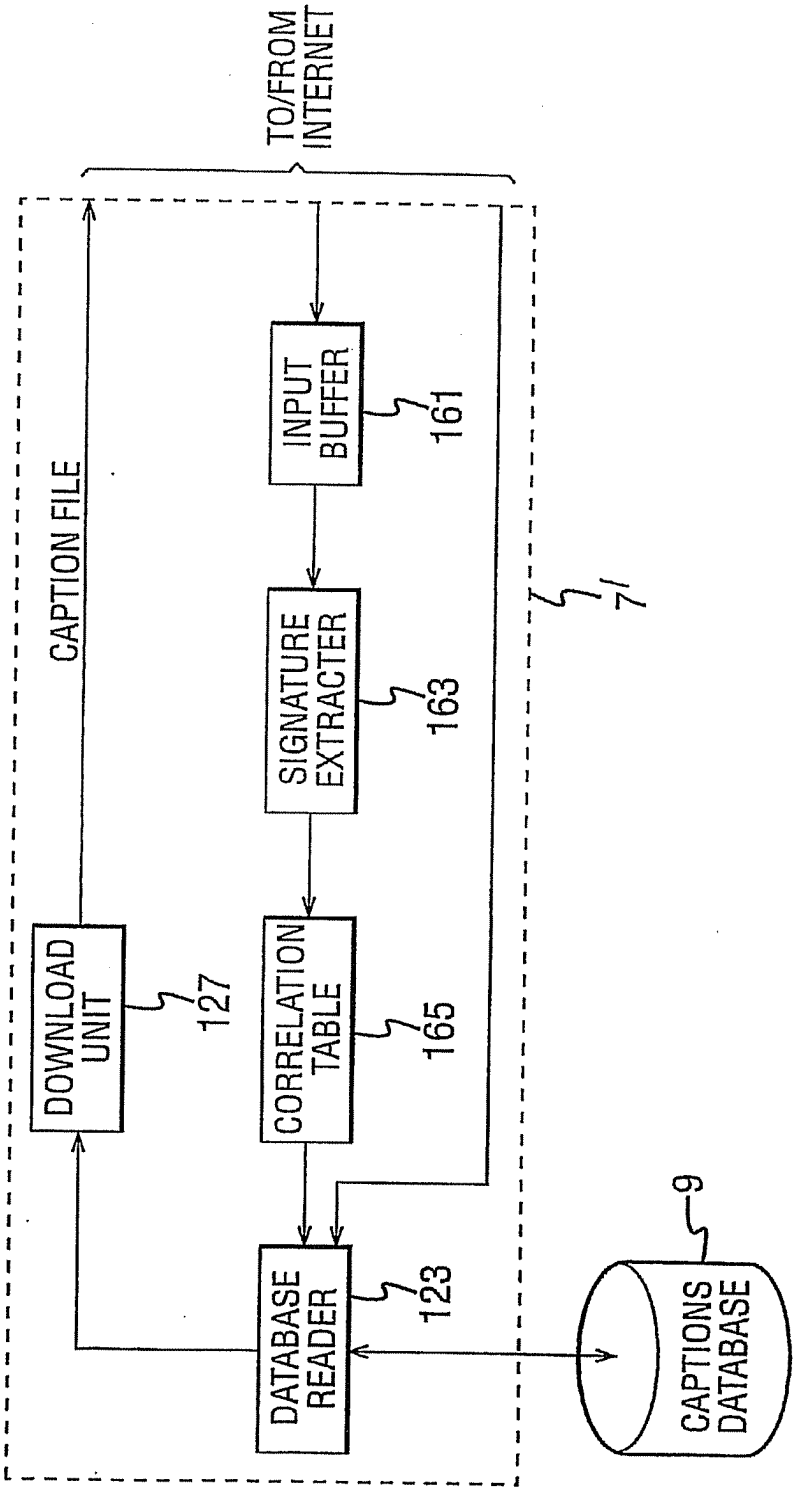

TIMING CONTROLLER
85
CLK
87
DATA EXTRACTER
145
CONTROL UNIT
81
INPUT BUFFER
47
143
AUDIO CODER
141
CAPTION MEMORY
65
147
CAPTION PLAYER
ADC
45
DAC
57
ADC
63
DAC
149
FILTER
43
TRANSMISSION CIRCUIT
55
RECEPTION CIRCUIT
61
AMPLIFIER
151
KEYPAD
97
41
53
153

FIG. 5


TO/FROM INTERNET
CAPTION FILE
INPUT BUFFER
161
SIGNATURE EXTRACTER
163
7
DOWNLOAD UNIT
127
CORRELATION TABLE
165
DATABASE READER
123
9
CAPTIONS DATABASE